ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis

Authors

Johannes Griss, Guilherme Viteri, Konstantinos Sidiropoulos, Vy Nguyen, Antonio Fabregat, and Henning Hermjakob

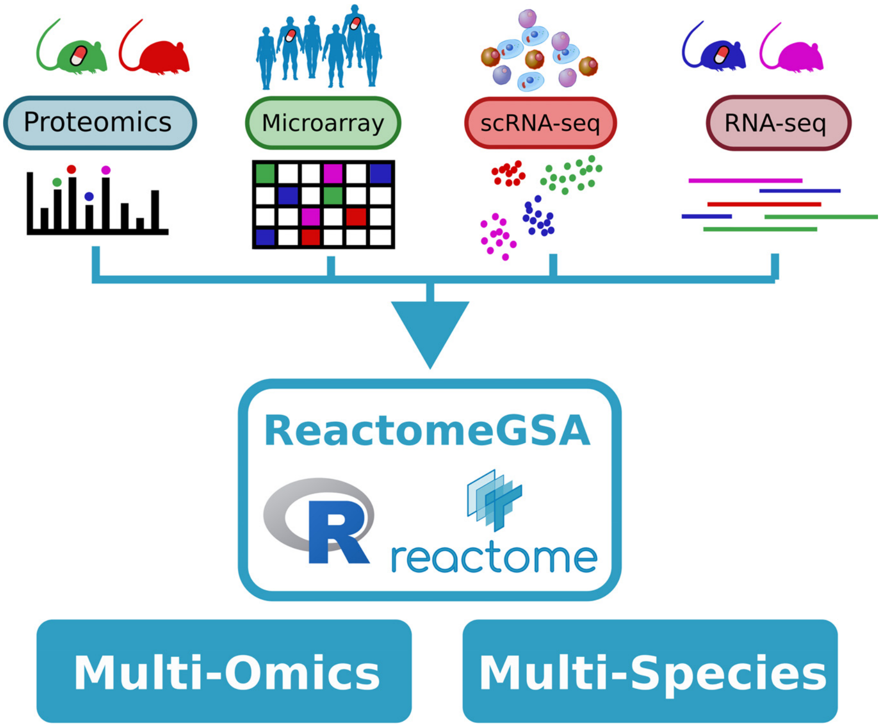
Correspondence

johannes.griss@meduniwien.ac.at;
hhe@ebi.ac.uk

In Brief

We present the novel ReactomeGSA resource for comparative pathway analyses of multi-omics datasets. ReactomeGSA is accessible through Reactome's web interface and the novel ReactomeGSA R Bioconductor package with explicit support for scRNA-seq data.We showcase ReactomeGSA's functionality by characterizing the role of B cells in anti-tumour immunity. Combining multi-omics data of five TCGA studies reveals marked opposing effects of B cells in different cancers. This showcases how ReactomeGSA can quickly derive novel biomedical insights by integrating large multi-omics datasets.

Graphical Abstract



Highlights

- ReactomeGSA is a novel tool for multi-species, multi-omics pathway analysis.
- Its quantitative pathway analysis methods offer high statistical power.
- Combining data of five TCGA studies shows B cells have opposing effects in cancers.
- ReactomeGSA reveals differences in key pathways between transcript- and protein-level.

Griss et al., 2020, Mol Cell Proteomics *19(12),* 2115–2124
December 2020 © 2020 Griss et al. Published by The American Society for Biochemistry and Molecular Biology, Inc.
https://doi.org/10.1074/mcp.TIR120.002155

MCP TECHNOLOGICAL INNOVATION AND RESOURCES

⌘ *Author's Choice*

ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis

Johannes Griss[1,2,*], Guilherme Viteri[1], Konstantinos Sidiropoulos[1], Vy Nguyen[2], Antonio Fabregat[1], and Henning Hermjakob[1,*]

Pathway analyses are key methods to analyze 'omics experiments. Nevertheless, integrating data from different 'omics technologies and different species still requires considerable bioinformatics knowledge.

Here we present the novel ReactomeGSA resource for comparative pathway analyses of multi-omics datasets. ReactomeGSA can be used through Reactome's existing web interface and the novel ReactomeGSA R Bioconductor package with explicit support for scRNA-seq data. Data from different species is automatically mapped to a common pathway space. Public data from ExpressionAtlas and Single Cell ExpressionAtlas can be directly integrated in the analysis. ReactomeGSA greatly reduces the technical barrier for multi-omics, cross-species, comparative pathway analyses.

We used ReactomeGSA to characterize the role of B cells in anti-tumor immunity. We compared B cell rich and poor human cancer samples from five of the Cancer Genome Atlas (TCGA) transcriptomics and two of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) proteomics studies. B cell-rich lung adenocarcinoma samples lacked the otherwise present activation through NFkappaB. This may be linked to the presence of a specific subset of tumor associated IgG+ plasma cells that lack NFkappaB activation in scRNA-seq data from human melanoma. This showcases how ReactomeGSA can derive novel biomedical insights by integrating large multi-omics datasets.

Increasingly available approaches such as transcriptome sequencing (RNA-seq), MS-based shotgun proteomics, and microarray studies enable us to characterize genome- and proteome-wide expression changes. This leads to the challenge of deriving relevant biological insights from lists of hundreds of regulated genes and proteins.

Pathway analysis techniques have emerged as a solution to this problem. Resources like the Gene Ontology (GO) (1), the Kyoto Encyclopedia of Genes and Genomes (KEGG) (2), the Molecular Signatures Database (MSigDB) (3), or Reactome (4) organize existing biological knowledge into gene sets or pathways. Pathway analysis approaches can use these resources to represent long lists of regulated genes and proteins as biologically defined pathways. This leads to a more intuitive interpretation of the data and increases the statistical power. Although single genes or proteins may only show small, nonsignificant changes, synchronous changes within a pathway may reveal a biologically important effect. Thereby, pathway analysis has become an essential resource for 'omics data analyses.

The increasing availability of public 'omics datasets has made it common practice to include these into analyses. These data integration is commonly complicated if datasets were created in different species or using different 'omics approaches. Pathway analysis approaches offer a solution to this problem because data can be mapped to the more general and comparable pathway space.

Existing web-based pathway analysis resources, such as PANTHER (5), the Database for Annotation, visualization and Integrated Discovery (DAVID) (6) or Reactome's pathway analysis (7) all provide over-representation analyses. This type of pathway analysis only tests whether a list of genes is overrepresented in a specific pathway. These approaches have the advantage that the user input is simple, but ignore any underlying quantitative information at the cost of reduced statistical power. Moreover, users must manually separate up- and down-regulated genes and process them in separate analyses. Thereby, any result is only a partial representation of the underlying biological changes.

The recently developed iLINCS resource extends the concept of single-resource pathway analysis to a powerful multi-omics and multi-resource analysis (8). It tests whether a list of gene/protein identifiers correlates with a large set of pre-computed signatures. These signatures are often the result of differential expression analyses. Therefore, like the aforementioned resources, iLINCS ignores any underlying quantitative information in the final comparison. Additionally, the comparison with public data are limited to pre-defined

From the [1]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridgeshire, United Kingdom; [2]Department of Dermatology, Medical University of Vienna, Vienna, Austria

This article contains supplemental data.

*For correspondence: Johannes Griss, johannes.griss@meduniwien.ac.at, Henning Hermjakob, hhe@ebi.ac.uk.

DOI 10.1074/mcp.TIR120.002155

experimental designs and comparisons whose results are stored as pre-computed signatures. Therefore, a large portion of the data remains unused.

Here, we present the novel Reactome gene set analysis system "ReactomeGSA." ReactomeGSA supports the comparative pathway analysis of multiple independent datasets. Datasets are submitted to a single pathway analysis and represented side-by-side on the pathway level. It uses gene set analysis methods that take the quantitative information into consideration and thereby performs the differential expression analysis directly on the pathway level. Data from different species is automatically mapped to a common pathway space through Reactome's internal mapping system. All supported gene set analysis methods are optimized for different types of 'omics approaches including single cell RNA-sequencing (scRNA-seq) data. Public datasets can be directly integrated from ExpressionAtlas and Single Cell ExpressionAtlas (9). We used ReactomeGSA to show that B cell receptor signaling is surprisingly down-regulated in B cell-rich lung adenocarcinoma in contrast to four other human cancers. We could further link this to IgG+ plasma cells in scRNA-seq data. ReactomeGSA thereby provides easy access to multi-omics, cross-species, comparative pathway analysis to reveal key biological mechanisms by integrating large 'omics datasets.

EXPERIMENTAL PROCEDURES

The ReactomeGSA analysis system is accessible through Reactome's web-based pathway browser application (https://www.reactome.org) and the "ReactomeGSA" R Bioconductor package. Both access ReactomeGSA's web-based application programming interface (API) which is also publicly accessible at https://gsa.reactome.org.

The backend is a Kubernetes application (https://kubernetes.io/) currently consisting of six deployments. Each deployment represents one Docker container (Docker Inc, https://www.docker.com). All data are stored in a Redis instance (https://redis.io/). The different components are linked through a message system provided by RabbitMQ (Pivotal, https://www.rabbitmq.com/). All components of the ReactomeGSA backend are developed in Python. The actual gene set analysis is performed using R Bioconductor (10) packages through the rpy2 (https://rpy2.github.io/) Python interface to the R language in the worker node (Fig. 1).

A key advantage of this setup is that the complete ReactomeGSA application can be described in one so-called YAML file - a Kubernetes configuration file. Because all Docker containers are freely available on Docker Hub (https://hub.docker.com) the ReactomeGSA system can be deployed using the single "kubectl apply -f reactome_gsa.yaml" command. We created a single YAML-formatted configuration file to quickly adapt ReactomeGSA to different use cases (ie. the number of resources available to the different nodes). Detailed information on how to adapt ReactomeGSA can be found on the GitHub repository (https://github.com/reactome/ReactomeGSA). Thereby, users can set up their own version of the ReactomeGSA system within minutes and deploy it locally or in the cloud.

Multi-Omics Gene Set Analysis—At the time of writing, ReactomeGSA supports three different analysis methods: Camera through the "limma" (11) package, PADOG through the "PADOG" package (12), and the single-sample gene set enrichment analysis (ssGSEA) (13) through the "GSVA" (14) package. All pathway analyses are performed by the worker node in the ReactomeGSA system (Fig. 1).

The workflow in ReactomeGSA follows the following briefly described steps: First, the user's input data are validated in terms of experimental design, validity of submitted identifiers, and data format. Next, all identifiers are mapped to the respective human UniProt identifiers (see below). Then, the selected pathway analysis is performed for each of the submitted datasets. The parameters for the pathway analysis (such as the kernel to use for the ssGSEA analysis) is automatically chosen based on the selected data type. Finally, the pathway analysis result is converted to Reactome's internal data format to render the result in the PathwayBrowser.

Reactome's manual curation is based on human UniProt identifiers (15). Thus, as a first step in the analysis, the submitted identifiers are mapped to human UniProt using Reactome's identifier mapping system. A key issue in mapping identifiers between different identifier systems and across species is to resolve one-to-many mappings. In these cases, the ReactomeGSA system keeps an internal record of all mappings. Genes that map to multiple UniProt identifiers which all belong to the same pathway are only added once to this pathway. Thereby, one-to-many mappings are resolved at the pathway-level and inaccuracies introduced through identifier conversions are greatly reduced.

To increase the coverage of Reactome pathways, pathways can be extended through medium and high confidence interactions derived from IntAct (16). This function considerably extends Reactome's coverage.

At the time of writing, the ReactomeGSA system supports five types of quantitative 'omics data: Microarray intensities, transcriptomics raw and normalized read counts, and proteomics spectral counts and intensity-based quantitative data. Internally, these different types of data are processed using two different methods: statistics for discrete quantitative data (in case of raw transcriptomics read counts and spectral counting based quantitative proteomics data) and statistics for continuous data. For Camera and PADOG, discrete values are normalized using edgeR's (17) calcNormFactors function. Then, the data are transformed using limma's voom function (18). Continuous data are directly processed using limma (11) and normalized using limma's normalizeBetweenArrays function. The pathway analysis is subsequently performed using limma's camera function or PADOG as implemented in the respective Bioconductor R package (19). For the ssGSEA method (13) the analysis is performed using the GSVA Bioconductor R package (14). Discrete data are processed using a poisson kernel and continuous data using a gaussian kernel. Thereby, multiple types of 'omics data can be supported.

scRNA-Seq Pathway Analysis—The analysis of scRNA-seq data are supported through the ReactomeGSA R package's "analyze_sc_clusters" function, as well as through the direct import of data from the Single Cell Expression Atlas (9). In both cases, we calculate the mean expression of genes within a cluster. For the R package, this is done through either "Seurat"'s (20) "AverageExpression" function, or through scater's (21) "aggregateAccrossCells" function depending on the input object. Single cell data retrieved from the Single Cell Expression Atlas is processed using custom python code (see https://github.com/reactome/gsa-backend for details). This approach to create pseudo-bulk RNA-seq data resembles previously described methods to calculate differentially expressed genes (22). Thereby, all pathway analysis methods supported by the ReactomeGSA analysis system are accessible to scRNA-seq data as well.

TCGA B Cell Analysis—The TCGA transcriptomics data for melanoma (TCGA-SKCM) (23), lung adenocarcinoma (TCGA-LUAD) (24), lung squamous cell carcinoma (TCGA-LUSC) (25), ovarian cancer (TCGA-OV) (26), and breast cancer (TCGA-BRCA) (27) were retrieved using the "TCGAbiolinks" R Bioconductor package (28). For all

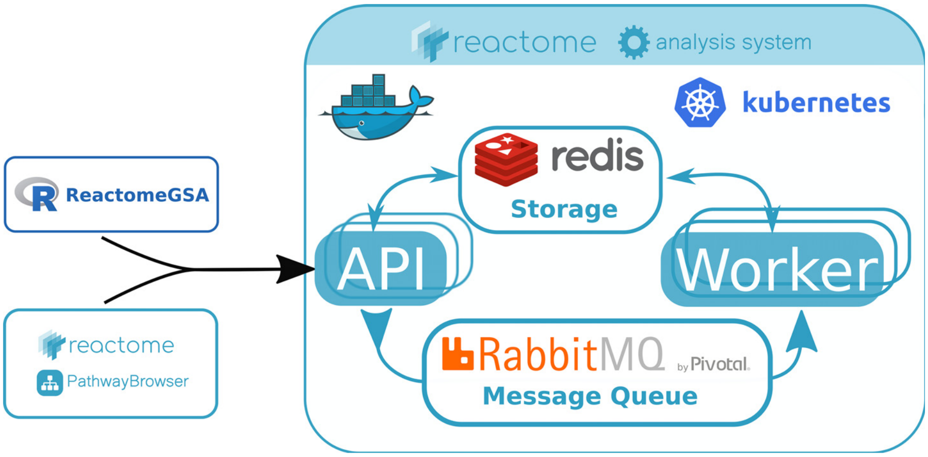


FIG. 1. **Schema of the ReactomeGSA system.** All requests are sent to a public web-based API through the ReactomeGSA Bioconductor R package or Reactome's web-based PathwayBrowser. The system is a Kubernetes application based on the microservices architecture. All requests are distributed through an internal message queue using RabbitMQ. Worker nodes are responsible for the complete pathway analysis, including identifier mapping and the creation of the visualization data in Reactome's pathway browser. Data nodes are responsible to load data from external resources such as ExpressionAtlas. Finally, report nodes create PDF and Microsoft Excel files as a static report of the analysis results. All data are stored in a central Redis instance. All nodes are Docker containers that are orchestrated by Kubernetes and automatically scaled based on current demand. Thereby, the application can dynamically adapt to changing usage levels.

datasets apart from melanoma, only primary tumor samples were retained. Genes that were expressed in less than 30% of the samples with at least 10 reads were removed.

The abundance of plasmablast-like B cells (TIPB) was quantified using the single-sample Gene Set Enrichment Analysis (ssGSEA) method (13) as implemented in the "GSVA" R Bioconductor package (14). Plasmablast-like B cells were described as CD38, CD27, and PAX5 (29). Samples were classified as TIPB-high and -low split by the median expression of the TIPB signature in all samples of the cohort. Overall survival was assessed using the R "survival" package.

The comparative pathway analysis was performed using the ReactomeGSA R Bioconductor package. In all studies, plasmablast "high" and "low" samples were compared with each other using PADOG (12).

The complete R code of this analysis, including the detailed versions of all R packages used is available in the respective Jupyter notebook (see Data availability).

CPTAC Data Analysis—Data processed through the common data analysis pipeline (CDA) was downloaded from the CPTAC data portal (breast cancer at https://cptac-data-portal.georgetown.edu/cptac/s/S015, ovarian cancer at https://cptac-data-portal.georgetown.edu/cptac/s/S020). For breast cancer (30), we used the proteome-level iTRAQ summary, for ovarian cancer (31) the PNNL-based protein-level iTRAQ summary. Samples were matched to the respective TCGA samples through the short barcode using the first 11 characters. Only unambiguous matches were retained. Plasmablast abundance-based groupings were transferred from the respective TCGA data set. The data were analyzed using the ReactomeGSA R package and PADOG.

Example scRNA-Seq Analysis—Raw read counts of the scRNA-seq data set by Jerby-Arnon *et al.* (32) were retrieved from the Gene Expression Omnibus (GEO, identifier GSE115978). The data were processed using "Seurat" version 3.1 (20) following the new scTransform normalization strategy regressing out the patient and cohort properties. To identify the B cells from the total number of cells we used the first 35 components of the principal component analysis for the subsequent steps. The neighbor graph and clustering was performed using the default parameters. B cell clusters were identified based on a high expression of CD20 (MS4A1), CD79A, CD19, and CD138 (SDC1).

B cells were extracted from the data set and re-processed, starting with the normalization step. Here, the top 11 components of the principal component analysis were used for the respective analysis steps. B cell clusters were subsequently classified following the strategy by Sanz *et al.* (33). Plasmablast-like B cells and plasma cells were differentiated based on a low expression of MS4A1 (CD20) in plasmablast-like B cells. Finally, the ssGSEA analysis was performed using the ReactomeGSA R packages' analyze_sc_clusters function.

The complete workflow including the detailed versions of all used R packages can be found in the respective Jupyter notebook (see Data availability).

RESULTS

ReactomeGSA can be accessed through Reactome's web interface (https://www.reactome.org/PathwayBrowser/#TOOL=AT) or through the novel "ReactomeGSA" R Bioconductor package (https://doi.org/doi:10.18129/B9.bioc.ReactomeGSA, Fig. 1). Both access the public API (https://gsa.reactome.org) to perform the pathway analysis. The analysis system is a Kubernetes application based on the microservice paradigm that automatically scales to current demand (see Methods for details). This infrastructure enables us to offer computationally expensive pathway analysis methods through an open interface. ReactomeGSA currently supports three methods: PADOG (12), Camera through the limma R package (11), and the ssGSEA (13) through the GSVA (14) R package (see Experimental Procedures for details). Although PADOG more often ranks biologically important pathways higher than other approaches, it is computationally more expensive. In such cases, Camera, which does not rely on permutations but linear models, results in faster results. ssGSEA is not a gene set enrichment analysis but aggregates expression values on the pathway level. This is helpful if the analyzed samples cannot be attributed to clear phenotypes or are to be correlated with continuous parameters such as survival time. The API and its complete specification is publicly available at https://gsa.reactome.org. Thereby, ReactomeGSA can easily be integrated into any other software infrastructure.

ReactomeGSA is fully integrated in Reactome's existing web-based pathway browser application (Fig. 2). After choosing the new "Analyse gene expression" tab and the desired analysis method, the user can add any number of datasets to the analysis request. Public datasets are directly loaded from Expression Atlas and the Single Cell Expression Atlas (9). Results can be sent as emails including static PDF and Microsoft Excel reports. Finally, the complete gene set analysis result is visualized in Reactome's interactive pathway browser. The pathway browser enables users to view Reactome's complete pathways from a tree-based, hierarchical overview, down to the single gene- and protein-level reactions. The results of different datasets can be switched at the click of a button or automatically changed every few seconds like a slideshow across all results. Thereby, differences between the analyzed datasets are immediately visible and can subsequently be interactively investigated down to the single gene or protein level.

The ReactomeGSA R package has been included in Bioconductor since version 3.10 (Fig. 3). Like the web interface, multiple datasets can be added to a ReactomeAnalysisRequest object. Expression values and metadata can directly be loaded from Bioconductor ExpressionSet, limma EList (11) and edgeR (17) DGEList objects. Thereby, the ReactomeGSA package can easily be integrated into existing R-based workflows. The analysis results are returned as a ReactomeAnalysisResult object. This object contains the pathway analysis results across all analyzed datasets, as well as the gene- or protein-level results of the differential expression analysis. It can directly open the interactive visualization in Reactome's web-based pathway browser (see above) and create plots to visualize the comparative pathway analysis results. Thereby, the multi-data set results generated by ReactomeGSA can be natively processed in R.

The ReactomeGSA R package has dedicated features to simplify pathway analyses of scRNA-seq data (Fig. 3). The "analyse_sc_clusters" function can directly process Seurat (20) and Bioconductor's SingleCellExperiment objects (22). It automatically retrieves the average gene expression per cell cluster and performs an ssGSEA analysis on the cluster-level expression values. This results in one pathway-level expression value per cell cluster. Thereby, cell clusters can quickly be interpreted based on specific biological functions.

ReactomeGSA Reveals a Lack of B Cell Activation in B Cell-Rich Lung Adenocarcinoma—We were among the first to show that B cells play a crucial role in anti-tumor immunity in human melanoma (29). *In vitro*, B cells differentiate toward a TIPB phenotype in the presence of melanoma cells. The corresponding molecular TIPB signature predicts overall survival in the TCGA melanoma cohort. Whether this effect is specific to melanoma or whether it is a general part of the anti-tumor immune response is currently unknown.

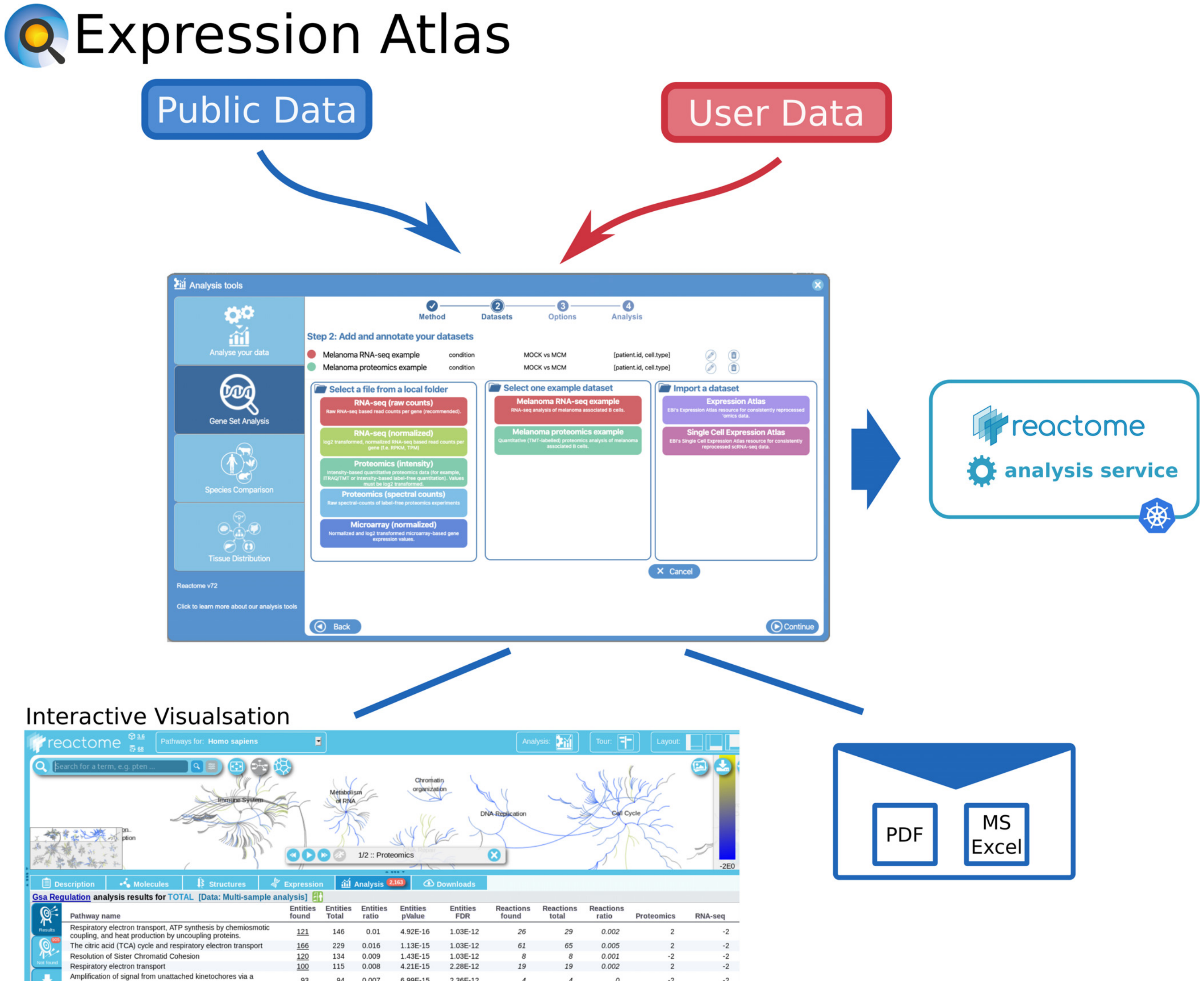
We analyzed the difference between TIPB-high *versus* TIPB-low samples in the TCGA cohorts for melanoma (23), lung adenocarcinoma (24), lung squamous cell carcinoma (25), ovarian cancer (26), and breast cancer (27). Melanoma and ovarian cancer patients with high levels of TIPB showed significantly longer overall survival (likelihood ratio test $p < 0.01$ for both, hazard ratio 0.56 melanoma, 0.69 ovarian cancer, Fig. 4*A*). There was no significant difference in overall survival for lung adenocarcinoma, lung squamous cell, and breast cancer patients (likelihood ratio test $p = 0.04$, $p = 0.2$ and $p = 0.9$ respectively). Therefore, the effect of TIPB on anti-tumor immunity and patient survival differs across these types of cancers.

We subsequently assessed pathway-level differences between patients with high- and low-levels of TIPB in the five cohorts. The comparative pathway analysis was performed using our ReactomeGSA R package and the PADOG gene set enrichment analysis. 383 pathways were significantly regulated in at least one of the datasets (FDR < 0.1, supplemental Data S1). 64 of these pathways showed a differential regulation in one of the datasets compared with melanoma. We previously showed *in vitro* that NF-kappaB activation was significantly up-regulated in B cells after stimulation with melanoma conditioned medium (29). Lung adenocarcinoma samples were the only ones that showed a significant down-regulation of the "Activation of NF-kappaB in B cells" pathway (FDR = 0.08). Even though these samples have a higher number of TIPB, overall B cell activation is reduced.

We specifically assessed how the lung adenocarcinoma cohort differs from the melanoma cohort. In total, 18 pathways were significantly regulated in both the melanoma and the lung adenocarcinoma cohort (Fig. 4*B*). Next to the down-regulation of NF-kappaB related genes, there was an overall down-regulation of B cell receptor signaling, but also p53 related DNA damage response, cell cycle and apoptosis related pathways. This shows that lung adenocarcinoma samples with a high number of tumor induced plasmablast-like B cells have a distinct different signaling state compared with melanoma.

Pathways related to B cell receptor signaling and apoptosis correlate with the survival benefit observed through higher numbers of TIPB. The melanoma and ovarian cancer cohort both showed the strongest survival benefit which was linked to the strongest up-regulation of apoptosis related pathways but also B cell receptor signaling. These results highlight that ReactomeGSA's comparative pathway analysis can quickly reveal clinically relevant conserved signaling events.

Cancer-Relevant Pathways Differ in Proteomics and Transcriptomics Data—In our recent characterization of melanoma



Step 2: Add and annotate your datasets

Melanoma RNA-seq example	condition	MOCK vs MCM	[patient.id, cell.type]
Melanoma proteomics example	condition	MOCK vs MCM	[patient.id, cell.type]

RNA-seq (raw counts): Raw RNA-seq based read counts per gene (recommended).

RNA-seq (normalized): log2 transformed, normalized RNA-seq based read counts per gene (f.e. RPKM, TPM)

Proteomics (intensity): Intensity-based quantitative proteomics data (for example, iTRAQ/TMT or intensity-based label-free quantitation). Values must be log2 transformed.

Proteomics (spectral counts): Raw spectral-counts of label-free proteomics experiments

Microarray (normalized): Normalized and log2 transformed microarray-based gene expression values.

Melanoma RNA-seq example: RNA-seq analysis of melanoma associated B cells.

Melanoma proteomics example: Quantitative (TMT-labelled) proteomics analysis of melanoma associated B cells.

Expression Atlas: EBI's Expression Atlas resource for consistently reprocessed 'omics data.

Single Cell Expression Atlas: EBI's Single Cell Expression Atlas resource for consistently reprocessed scRNA-seq data.

Gsa Regulation analysis results for TOTAL [Data: Multi-sample analysis]

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Proteomics	RNA-seq
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	121	146	0.01	4.92E-16	1.03E-12	26	29	0.002	2	-2
The citric acid (TCA) cycle and respiratory electron transport	166	229	0.016	1.13E-15	1.03E-12	61	65	0.005	2	-2
Resolution of Sister Chromatid Cohesion	120	134	0.009	1.43E-15	1.03E-12	8	8	0.001	-2	-2
Respiratory electron transport	100	115	0.008	4.21E-15	2.28E-12	19	19	0.002	2	-2
Amplification of signal from unattached kinetochores via a	93	94	0.007	6.99E-15	2.36E-12	4	4	0	-2	-2

FIG. 2. **ReactomeGSA is fully integrated into the web-based Reactome pathway browser (https://reactome.org).** Users can either upload their own datasets or import public data from ExpressionAtlas. The gene set analysis is performed through the ReactomeGSA API. Results are visualized in Reactome's interactive pathway browser and sent as static reports in PDF and Microsoft Excel format via E-mail.

associated B cells, key phenotypic changes in B cells were primarily observed on the protein but not the transcriptome level. We performed a comparative pathway analysis of the two TCGA cohorts that were also analyzed by CPTAC using a global proteomics approach.

99 samples of the breast cancer CPTAC study (30) and 62 samples of the CPTAC ovarian study (31) could be directly mapped to samples from the respective TCGA study. As our TIPB signature was only validated for transcriptomics data, sample grouping into TIPB-high and -low samples was transferred from the TCGA data. The pathway analysis was performed using our ReactomeGSA R package and PADOG. 113 and 96 pathways were significantly regulated (FDR $<$ 0.05, supplemental Data S2) in the proteomics and transcriptomics data from the breast and ovarian cancer study respectively. Out of these, 13 showed a different direction of regulation in the breast study, and one in the ovarian cancer study between proteomics and transcriptomics measurements. In breast cancer, these included VEGF signaling, EGFR signaling, and IGF1R signaling related pathways (all up-regulated in transcriptomics and down-regulated in proteomics). In ovarian cancer, FGFR signaling was significantly up-regulated in the transcriptomics but down-regulated in proteomics data. All of these pathways are linked to proliferation and are relevant pathways to tumor biology. B cell receptor signaling associated pathways were significantly up-regulated in all datasets. This highlights how ReactomeGSA can quickly reveal biologically relevant differences and similarities between ‘omics datasets.

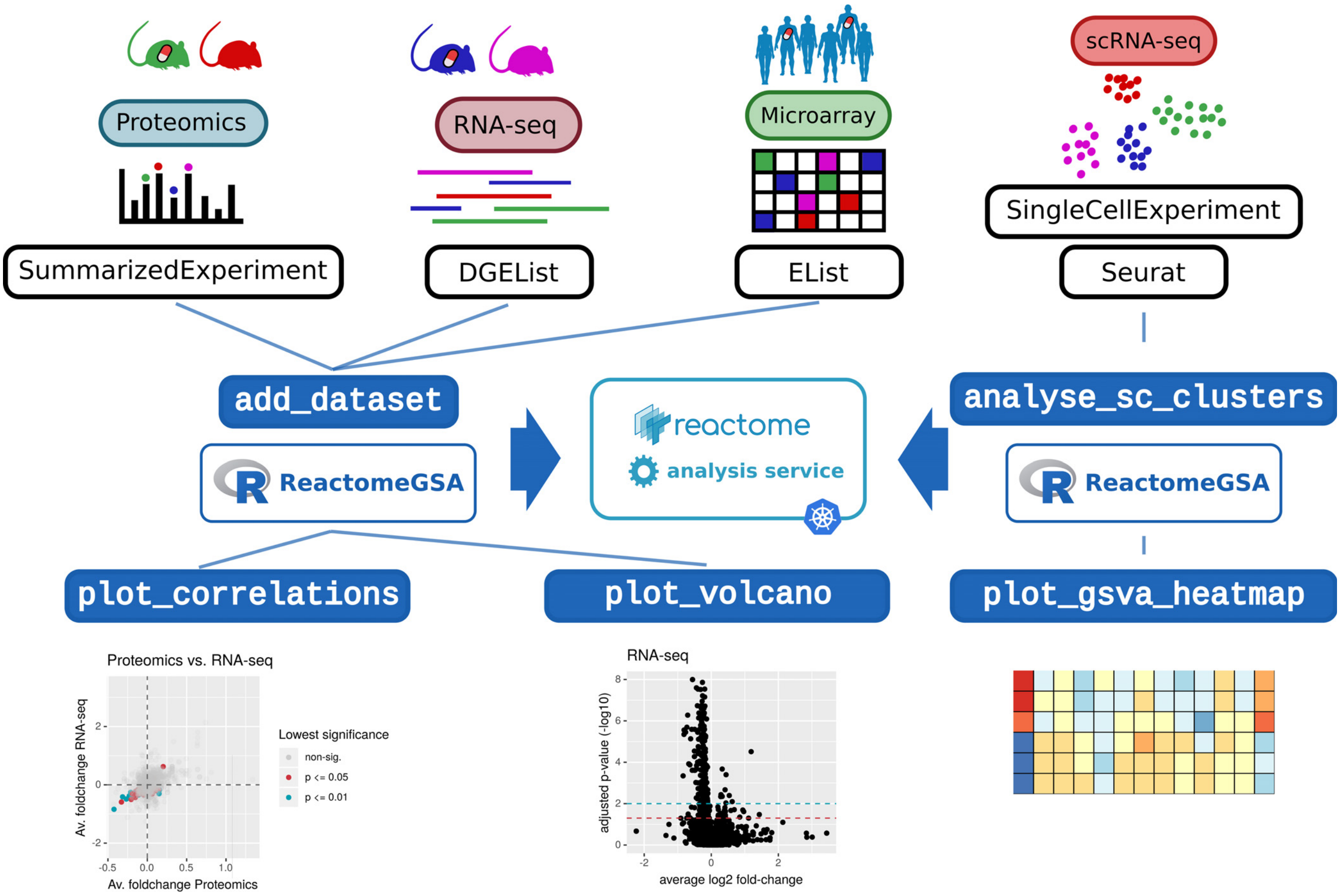


FIG. 3. **The ReactomeGSA Bioconductor R package can directly process data from the most commonly used data structures for 'omics analyses.** The pathway analysis is performed through the ReactomeGSA analysis system and made available through a native R object. Convenient plotting functions give a quick overview of how well two datasets correlate on the pathway level. Volcano plots further highlight the magnitude of the observed changes in individual datasets. Additionally, pathway analysis of scRNA-seq data are simplified through the single "analyse_sc_clusters" function.

IgG+ Plasma Cells Show Reduced NFkappaB Activation—Specific subtypes of B cells seem to be primarily responsible for the B cell triggered anti-tumor response (29, 34–36). We therefore assessed whether the observed difference in NFkappaB activation is B cell subtype specific.

The extracted B cells from the scRNA-seq data set by Jerby-Arnon *et al.* (32) formed 13 distinct clusters using Seurat (see Methods for details). Based on canonical B cell markers (33) we classified these clusters as double negative B cells, seven types of memory-like B cells, memory-switched resting and -activated B cells, naive B cells, plasma cells, and plasmablast-like B cells (Fig. 5*A*). Consistent with their transitional phenotype between B cells and plasma cells, plasmablast-like B cells were the only to express SDC1 (CD138) and low levels of MS4A1 (CD20). This classification already highlights issues in classifying B cell subtypes as we had to classify seven clusters as memory B cells even though they showed marked differences in overall gene expression.

We used ReactomeGSA R package's analyse_sc_clusters function to quantify pathways in these B cell clusters. There was a considerable heterogeneity between the memory B cell clusters, as well as plasmablast and plasma cells in terms of B cell receptor signaling (Fig. 5*B*). In the latter, this matches the previously described lack of functional B cell receptors in IgG positive plasma cells (37). Consistently, plasma cells but not plasmablast-like B cells expressed high levels of IgG as determined through Fc fragment of IgG receptor and transporter (FCGRT) expression (Fig. 5*C*). Plasma cells and plasmablast-like B cells further differed in NTRK signaling which regulates cell survival, proliferation and motility (38). Our original TIPB signature is too coarse to perfectly differentiate between plasma cells and plasmablast-like B cells. Therefore, the lack of B cell receptor signaling in lung adenocarcinoma samples points toward the high abundance of IgG+ plasma cells. These were shown to be negative prognostic factors in lung adenocarcinoma (39) which may explain the reduced survival benefit of TIPB there.

DISCUSSION

ReactomeGSA greatly decreases the technical challenge to perform pathway analyses of unrelated datasets irrespective

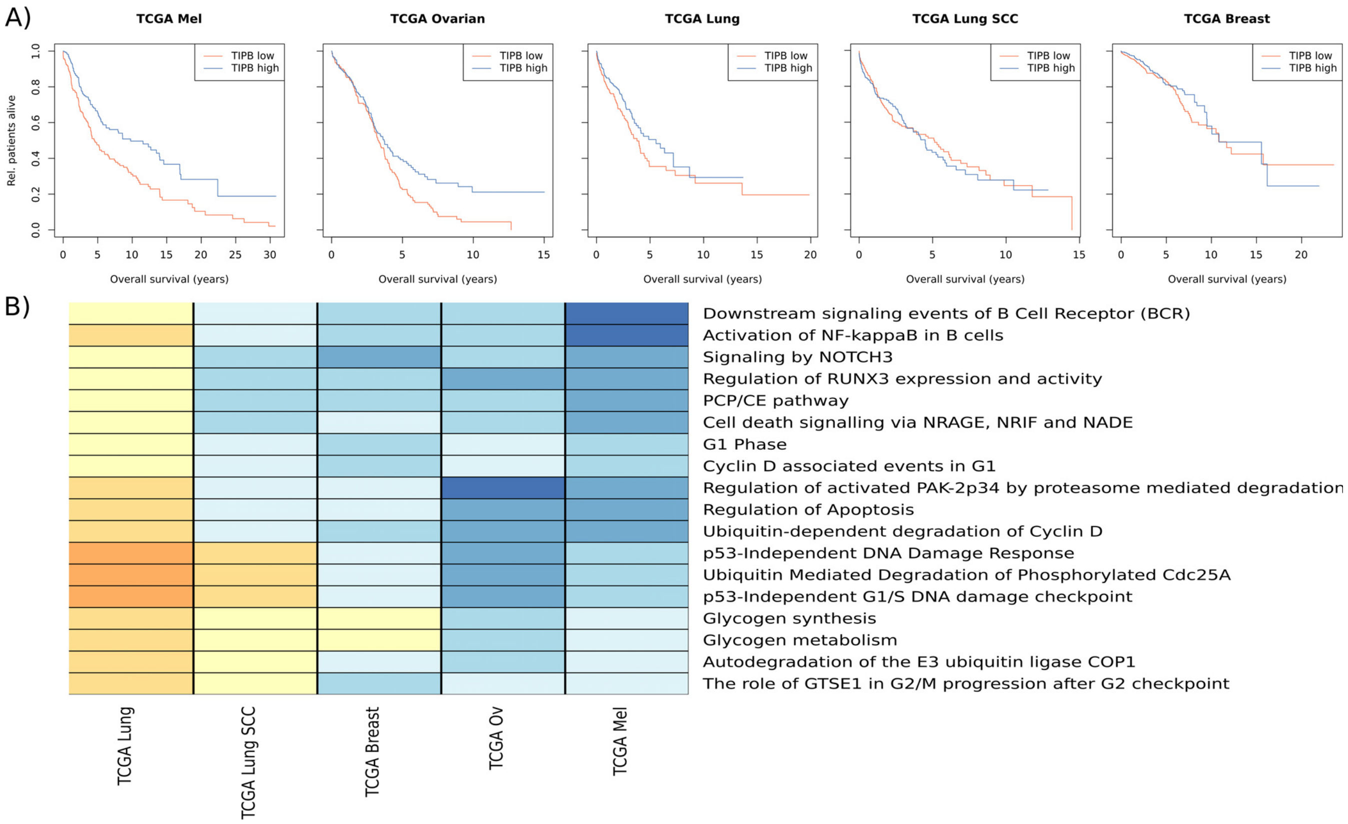


FIG. 4. **Comparison of TIPB-high *versus* -low samples from TCGA studies on melanoma (TCGA Mel), ovarian cancer (TCGA Ovarian), lung adenocarcinoma study (TCGA Lung), lung squamous cell carcinoma (TCGA Lung SCC), and breast cancer (TCGA Breast).** *A*, Overall survival of patients with high (*blue* line) or low (*red* line) expression of the TIPB signature (split by the median expression in the data set). *B*, Average gene fold-changes per pathway. Only pathways significantly regulated (FDR < 0.1) in the TCGA melanoma and the TCGA lung adenocarcinoma cohort with a different direction of regulation in these two cohorts are shown. Shades of yellow represent a down-regulation, shades of blue an up-regulation.

of ‘omics technology and investigated species. The iLINCS resource (8) is comparable in terms of the integration of different ‘omics data types and public datasets. In contrast to iLINCS, ReactomeGSA does not rely on pre-computed signatures for public datasets. This limits the number of public datasets that can be integrated into a single analysis. At the same time, it gives the researcher complete freedom in terms of experimental design and data analysis strategy to use. Our analysis of TCGA datasets based on a custom signature, for example, would not be supported by iLINCS. Additionally, ReactomeGSA directly supports quantitative ‘omics data as input. Thereby, we can use gene set analysis approaches with increased statistical power compared with simple overrepresentation analysis (19). The support for sample-level quantitative data enables us to integrate gene set variation analyses which we found especially helpful in the analysis of scRNA-seq data. We, thus, believe that the ReactomeGSA system is a considerable step forward in giving researchers easy access to complex, more sophisticated pathway analysis methods.

Nevertheless, ReactomeGSA is still limited to three “classic” ‘omics technologies. Future plans involve supporting methods such as chromatin accessibility sequencing data. Internally, ReactomeGSA is already designed to handle different types of quantitative data. ReactomeGSA thus provides an infrastructure that is well suited to cover a large variety of ‘omics technologies.

A key decision in multi-omics pathway analyses is how to integrate different types of ‘omics data. Methods such as the Gene Set Omic Analysis (GSOA) (40) or the PAthway Recognition Algorithm using Data Integration on Genomic Models (PARADIGM) (41, 42) merge different ‘omics measurements into a single result. Thereby, only data from the same or highly similar samples can be integrated. Moreover, differences between the different ‘omics measurements disappear. As highlighted in our example data and previous studies, such differences are to be expected (29, 30). We deliberately developed a system that can highlight such differences that researchers can interactively investigate with the Reactome pathway browser. Moreover, the user can quickly choose

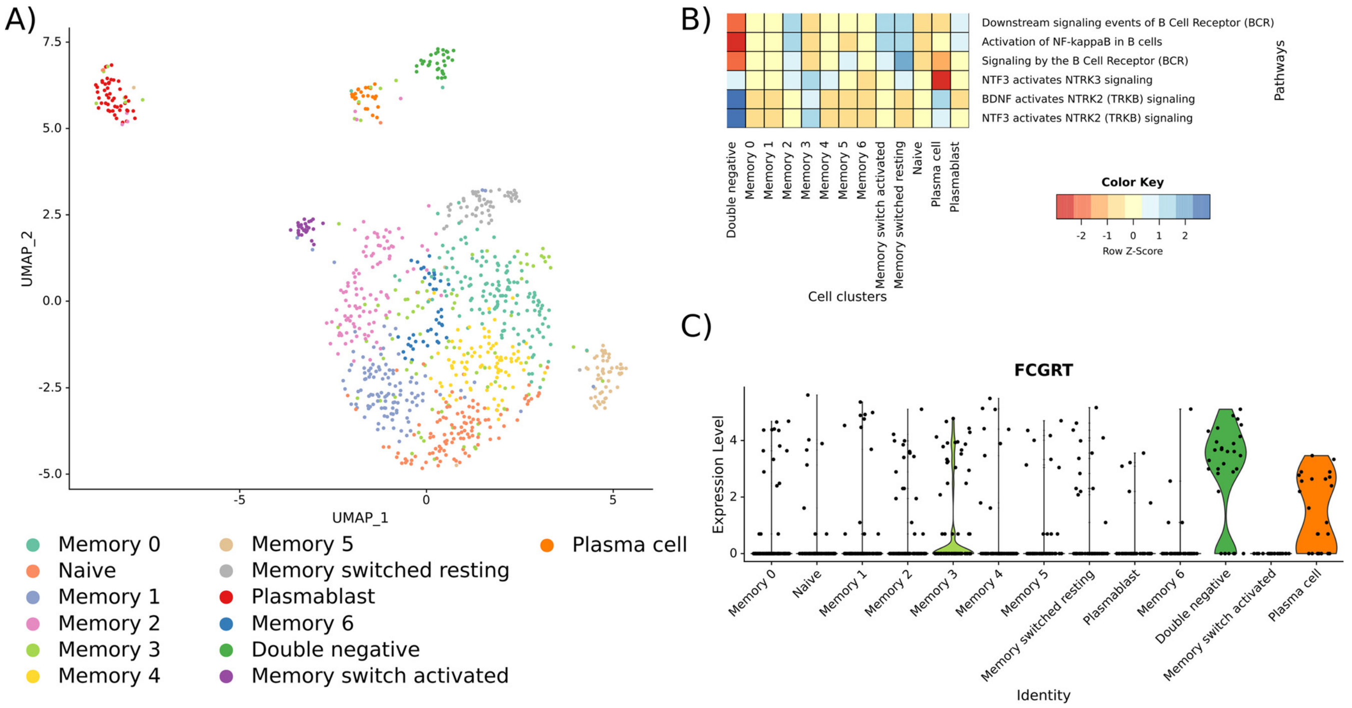


FIG. 5. **Analysis of B cell subtypes from the data set by Jerby-Arnon *et al.* (32)** *A*, UMAP plot of the identified B cell clusters. Cell type annotations are based on canonical B cell markers (33). *B*, ReactomeGSA gene set variation based pathway-level expression in the identified B cell clusters of the Jerby-Arnon *et al.* Data set. Expression values were z-score normalized by pathway. *C*, Expression of IgG estimated through FCGRT abundance in the B cell clusters.

between different pathway analysis algorithms that all have different strengths and weaknesses (43). ReactomeGSA provides a novel multi-omics pathway analysis infrastructure that is tailored to expert bioinformaticians and nonexperts alike.

DATA AVAILABILITY

The complete source code of the ReactomeGSA backend, the web-based pathway browser, and the ReactomeGSA Bioconductor R package are available under a permissive open source license on GitHub (https://github.com/reactome). All docker images of the ReactomeGSA analysis system are publically available on Docker Hub (https://hub.docker.com). Central links to all components of the ReactomeGSA system can be found at https://reactome.github.io/ReactomeGSA. The source code of the backend (ie. the Kubernetes application) can be found at https://github.com/reactome/gsa-backend. The source code of the R package is available at https://github.com/reactome/ReactomeGSA. Additionally, a detailed documentation on how to set up the ReactomeGSA analysis system on a local Kubernetes instance can be found on https://reactome.github.io/ReactomeGSA.

The detailed API specification of the ReactomeGSA system is available on https://gsa.reactome.org. Therefore, the complete analysis capabilities can easily be integrated into any other existing software platform.

The code to analyze the example datasets presented in this manuscript can be found as Jupyter notebooks on https://github.com/Reactome/ReactomeGSA-tutorials.

Funding and additional information—This project received funding from the Research Executive Agency (REA) under the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 788042, the US National Institutes of Health (P41 HG003751), and the European Molecular Biology Laboratory. This work was further supported by the FWF-Austrian Science Fund (project P30325-B28).

Author contributions—JG and HH designed research; JG, GV, KS, and VN performed research; JG, GV, KS, VN, and AF contributed new reagents/analytic tools; JG analyzed data; JG and HH wrote the paper.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are:API, application programming interface; CDA, common data analysis pipeline; CPTAC, Clinical Proteomic Tumor Analysis Consortium; DAVID, Database for Annotation, visualization and Integrated Discovery; GSOA, gene set omic analysis; GEO, gene expression omnibus; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MSigDB, Molecular Signatures

Database; PARADIGM, pathway recognition algorithm using data integration on genomic models; RNA-seq, transcriptome sequencing; scRNA-seq, single cell RNA-sequencing; ssGSEA, single-sample gene set enrichment analysis; TCGA, The Cancer Genome Atlas; TIPB, tumor induced plasmablast-like B cells.

Received June 2, 2020, and in revised form, July 28, 2020 Published, MCP Papers in Press, September 9, 2020, DOI 10.1074/mcp.TIR120.002155

REFERENCES

1. The Gene Ontology Consortium, (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. **47,** D330–D338
2. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. **45,** D353–D361
3. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A*. **102,** 15545–15550
4. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., and D'Eustachio, P. (2020) The reactome pathway knowledgebase. *Nucleic Acids Res*. **48,** D498–D503
5. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P. D. (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. **45,** D183–D189
6. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc*. **4,** 44–57
7. Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., D'Eustachio, P., Stein, L., and Hermjakob, H. (2017) Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics*. **18,** 142
8. Pilarczyk, M., Najafabadi, M. F., Kouril, M., Vasiliauskas, J., Niu, W., Shamsaei, B., Mahi, N., Zhang, L., Clark, N., Ren, Y., White, S., Karim, R., Xu, H., Biesiada, J., Bennet, M. F., Davidson, S., Reichard, J. F., Stathias, V., Koleti, A., Vidovic, D., Clark, D. J. B., Schurer, S., Ma'ayan, A., Meller, J., and Medvedovic, M. (2019) Connecting omics signatures of diseases, drugs, and mechanisms of actions with iLINCS. *bioRxiv 826271v1* **13**
9. Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A. M.-P., George, N., Fexova, S., Fonseca, N. A., Füllgrabe, A., Green, M., Huang, N., Huerta, L., Iqbal, H., Jianu, M., Mohammed, S., Zhao, L., Jarnuczak, A. F., Jupp, S., Marioni, J., Meyer, K., Petryszak, R., Prada Medina, C. A., Talavera-López, C., Teichmann, S., Vizcaino, J. A., and Brazma, A. (2020) Expression Atlas update: from tissues to single cells. *Nucleic Acids Res*. **48,** D77–D83
10. Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., Pagès, H., Reyes, A., Shannon, P., Smyth, G. K., Tenenbaum, D., Waldron, L., and Morgan, M. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12,** 115–121
11. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43,** e47
12. Tarca, A. L., Draghici, S., Bhatti, G., and Romero, R. (2012) Down-weighting overlapping genes improves gene set analysis. *BMC Bioinformatics*. **13,** 136
13. Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E. M., Sos, M. L., Michel, K., Mermel, C., Silver, S. J., Weir, B. A., Reiling, J. H., Sheng, Q., Gupta, P. B., Wadlow, R. C., Le, H., Hoersch, S., Wittner, B. S., Ramaswamy, S., Livingston, D. M., Sabatini, D. M., Meyerson, M., Thomas, R. K., Lander, E. S., Mesirov, J. P., Root, D. E., Gilliland, D. G., Jacks, T., and Hahn, W. C. (2009) Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462,** 108–112
14. Hänzelmann, S., Castelo, R., and Guinney, J. (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. **14,** 7
15. UniProt Consortium, (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. **47,** D506–D515
16. Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N. H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R. C., Meldal, B., Melidoni, A. N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G., and Hermjakob, H. (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*. **42,** D358–D363
17. McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. **40,** 4288–4297
18. Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. **15,** R29
19. Tarca, A. L., Bhatti, G., and Romero, R. (2013) A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity. *PLoS ONE*. **8,** e79217
20. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019) Comprehensive Integration of Single-Cell Data. *Cell* **177,** 1888–1902.e21
21. McCarthy, D. J., Campbell, K. R., Lun, A. T. L., and Wills, Q. F. (2017) Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33,** 1179–1186
22. Amezquita, R. A., Lun, A. T. L., Becht, E., Carey, V. J., Carpp, L. N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Soneson, C., Waldron, L., Pagès, H., Smith, M. L., Huber, W., Morgan, M., Gottardo, R., and Hicks, S. C. (2020) Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17,** 137–145
23. Cancer Genome Atlas Network, (2015) Genomic Classification of Cutaneous Melanoma. *Cell* **161,** 1681–1696
24. Cancer Genome Atlas Research Network, (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550
25. Cancer Genome Atlas Research Network, (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525
26. Cancer Genome Atlas Research Network, (2011) Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615
27. Cancer Genome Atlas Network, (2012) Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70
28. Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T. S., Malta, T. M., Pagnotta, S. M., Castiglioni, I., Ceccarelli, M., Bontempi, G., and Noushmehr, H. (2016) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. **44,** e71–e71
29. Griss, J., Bauer, W., Wagner, C., Simon, M., Chen, M., Grabmeier-Pfistershammer, K., Maurer-Granofszky, M., Roka, F., Penz, T., Bock, C., Zhang, G., Herlyn, M., Glatz, K., Läubli, H., Mertz, K. D., Petzelbauer, P., Wiesner, T., Hartl, M., Pickl, W. F., Somasundaram, R., Steinberger, P., and Wagner, S. N. (2019) B cells sustain inflammation and predict response to immune checkpoint blockade in human melanoma. *Nat. Commun*. **10,** 4186
30. Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., Wang, X., Qiao, J. W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J. T., Gatza, M. L., Wilkerson, M., Perou, C. M., Yellapantula, V., Huang, K.-L., Lin, C., McLellan, M. D., Yan, P., Davies, S. R., Townsend, R. R., Skates, S. J., Wang, J., Zhang, B., Kinsinger, C. R., Mesri, M., Rodriguez, H., Ding, L., Paulovich, A. G., Fenyö, D., Ellis, M. J., and Carr, S. A. and NCI CPTAC, (2016) Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534,** 55–62
31. Zhang, H., Liu, T., Zhang, Z., Payne, S. H., Zhang, B., McDermott, J. E., Zhou, J.-Y., Petyuk, V. A., Chen, L., Ray, D., Sun, S., Yang, F., Chen, L., Wang, J., Shah, P., Cha, S. W., Aiyetan, P., Woo, S., Tian, Y., Gritsenko, M. A., Clauss, T. R., Choi, C., Monroe, M. E., Thomas, S., Nie, S., Wu, C., Moore, R. J., Yu, K.-H., Tabb, D. L., Fenyö, D., Bafna, V., Wang, Y., Rodriguez, H., Boja, E. S., Hiltke, T., Rivers, R. C., Sokoll, L., Zhu, H., Shih, I.-M., Cope, L., Pandey, A., Zhang, B., Snyder, M. P., Levine, D. A., Smith, R. D.,

Chan, D. W., and Rodland, K. D. and CPTAC Investigators, (2016) Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **166,** 755–765

32. Jerby-Arnon, L., Shah, P., Cuoco, M. S., Rodman, C., Su, M.-J., Melms, J. C., Leeson, R., Kanodia, A., Mei, S., Lin, J.-R., Wang, S., Rabasha, B., Liu, D., Zhang, G., Margolais, C., Ashenberg, O., Ott, P. A., Buchbinder, E. I., Haq, R., Hodi, F. S., Boland, G. M., Sullivan, R. J., Frederick, D. T., Miao, B., Moll, T., Flaherty, K. T., Herlyn, M., Jenkins, R. W., Thummalapalli, R., Kowalczyk, M. S., Cañadas, I., Schilling, B., Cartwright, A. N. R., Luoma, A. M., Malu, S., Hwu, P., Bernatchez, C., Forget, M.-A., Barbie, D. A., Shalek, A. K., Tirosh, I., Sorger, P. K., Wucherpfennig, K., Van Allen, E. M., Schadendorf, D., Johnson, B. E., Rotem, A., Rozenblatt-Rosen, O., Garraway, L. A., Yoon, C. H., Izar, B., and Regev, A, e24, (2018) A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175,** 984–997
33. Sanz, I., Wei, C., Jenks, S. A., Cashman, K. S., Tipton, C., Woodruff, M. C., Hom, J., and Lee, F. E.-H. (2019) Challenges and opportunities for consistent classification of human B cell and plasma cell populations. *Front. Immunol*. **10,** 2458
34. Cabrita, R., Lauss, M., Sanna, A., Donia, M., Skaarup Larsen, M., Mitra, S., Johansson, I., Phung, B., Harbst, K., Vallon-Christersson, J., van Schoiack, A., Lövgren, K., Warren, S., Jirström, K., Olsson, H., Pietras, K., Ingvar, C., Isaksson, K., Schadendorf, D., Schmidt, H., Bastholt, L., Carneiro, A., Wargo, J. A., Svane, I. M., and Jönsson, G. (2020) Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **577,** 561–565
35. Helmink, B. A., Reddy, S. M., Gao, J., Zhang, S., Basar, R., Thakur, R., Yizhak, K., Sade-Feldman, M., Blando, J., Han, G., Gopalakrishnan, V., Xi, Y., Zhao, H., Amaria, R. N., Tawbi, H. A., Cogdill, A. P., Liu, W., LeBleu, V. S., Kugeratski, F. G., Patel, S., Davies, M. A., Hwu, P., Lee, J. E., Gershenwald, J. E., Lucci, A., Arora, R., Woodman, S., Keung, E. Z., Gaudreau, P.-O., Reuben, A., Spencer, C. N., Burton, E. M., Haydu, L. E., Lazar, A. J., Zapassodi, R., Hudgens, C. W., Ledesma, D. A., Ong, S., Bailey, M., Warren, S., Rao, D., Krijgsman, O., Rozeman, E. A., Peeper, D., Blank, C. U., Schumacher, T. N., Butterfield, L. H., Zelazowska, M. A., McBride, K. M., Kalluri, R., Allison, J., Petitprez, F., Fridman, W. H., Sautès-Fridman, C., Hacohen, N., Rezvani, K., Sharma, P., Tetzlaff, M. T., Wang, L., and Wargo, J. A. (2020) B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* **577,** 549–555
36. Lu, Y., Zhao, Q., Liao, J.-Y., Song, E., Xia, Q., Pan, J., Li, Y., Li, J., Zhou, B., Ye, Y., Di, C., Yu, S., Zeng, Y., and Su, S. (2020) Complement signals determine opposite effects of B cells in chemotherapy-induced immunity. *Cell* **180,** 1081–1097.e24
37. Pinto, D., Montani, E., Bolli, M., Garavaglia, G., Sallusto, F., Lanzavecchia, A., and Jarrossay, D. (2013) A functional BCR in human IgA and IgM plasma cells. *Blood* **121,** 4110–4114
38. Gromnitza, S., Lepa, C., Weide, T., Schwab, A., Pavenstädt, H., and George, B. (2018) Tropomyosin-related kinase C (TrkC) enhances podocyte migration by ERK-mediated WAVE2 activation. *FASEB J*. **32,** 1665–1676
39. Kurebayashi, Y., Emoto, K., Hayashi, Y., Kamiyama, I., Ohtsuka, T., Asamura, H., and Sakamoto, M. (2016) Comprehensive Immune Profiling of Lung Adenocarcinomas Reveals Four Immunosubtypes with Plasma Cell Subtype a Negative Indicator. *Cancer Immunol. Res*. **4,** 234–247
40. MacNeil, S. M., Johnson, W. E., Li, D. Y., Piccolo, S. R., and Bild, A. H. (2015) Inferring pathway dysregulation in cancers from multiple types of omic data. *Genome Med*. **7,** 61
41. Vaske, C. J., Benz, S. C., Sanborn, J. Z., Earl, D., Szeto, C., Zhu, J., Haussler, D., and Stuart, J. M. (2010) Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26,** i237–i237
42. Sedgewick, A. J., Benz, S. C., Rabizadeh, S., Soon-Shiong, P., and Vaske, C. J. (2013) Learning subgroup-specific regulatory interactions and regulator independence with PARADIGM. *Bioinformatics* **29,** i62–i70
43. Geistlinger, L., Csaba, G., Santarelli, M., Ramos, M., Schiffer, L., Turaga, N., Law, C., Davis, S., Carey, V., Morgan, M., Zimmer, R., and Waldron, L. (2020) Toward a gold standard for benchmarking gene set enrichment analysis. *Brief. Bioinform*. bbz158